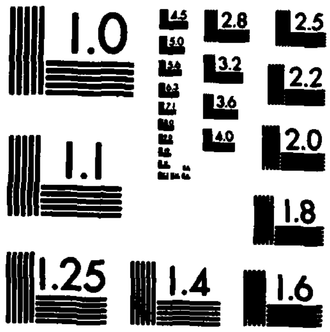
AD-A120 662 MODEL IDENTIFICATION AND ESTIMATION OF NONGAUSSIAN ARMA PROCESSES(U) CALIFORNIA UNIV SAN DIEGO LA JOLLA DEPT OF MATHEMATICS K LII 82 N00014-81-K-0003 1/1

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC


1.0
4.5
5.0
5.6
6.3
7.1
8.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

Model Identification and Estimation of NonGaussian ARMA Processes

By

Keh-Shin Lii

University of California, Riverside

Research supported in part by Office of Naval Research Contract N00014-81-K-0003

Key words: ARMA model, Identification, Estimation, C-table, Pade table, Asymptotics, Bispectrum.

AMS 1970 subject classifications: Primary 62M15; Secondary 62G05, 62E70.

Summary

Finite parameter models of ARMA type have been used extensively in many applications. Under the usual Gaussian assumption, the second order analysis will not be able to discriminate among competing models which give the same correlation structure. In many applications the innovation process is non-Gaussian. In this case, analysis using higher order moments will identify the model uniquely without the usual invertibility assumption. This in turn will affect the forecasting based on the non-Gaussian model. We present a method which uses bispectral analysis and the Pade approximation. We show that the method will consistently identify the order of the ARMA model and estimate the parameters of the model. One could also deconvolve the process to estimate the innovation process which will provide information for possible more efficient maximum likelihood estimation of the parameters. Asymptotic distributions are given, and a few examples are presented to illustrate the effectiveness of the method.

1.

1. Introduction

Finite parameter autoregressive moving average models have been used extensively in time series modeling, forecasting and control. Most of the literature is concerned with Gaussian processes. Let random variables e_t, $t = \ldots,-1,0,1,\ldots$ be independent and identically distributed with mean zero, $Ee_t = 0$, and variance one $Ee^2 = 1$. Let $\{a_j\}$ be a sequence of real constants with

$$\Sigma a_j^2 < \infty \quad .$$

Consider the linear process generated by $\{a_j\}$ and $\{e_t\}$

$$X_t = \sum_{j=-\infty}^{\infty} a_j e_{t-j} \quad . \tag{1.1}$$

The frequency response function is given by

$$A(e^{-i\lambda}) = \sum_j a_j e^{-ij\lambda} \quad . \tag{1.2}$$

If the process X_t is normally distributed then its full probability structure is completely determined by its spectral density function

$$f(\lambda) = \frac{1}{2\pi} \left|A(e^{-i\lambda})\right|^2 \quad . \tag{1.3}$$

Hence the phase information in $A(e^{-i\lambda})$ is not identifiable in the Gaussian case. If $A(z)$ is a rational function

$$A(z) = Q_q(z)/P_p(z) \tag{1.4}$$

with

$$\begin{aligned} Q_q(z) &= \sum_{i=0}^{q} q_i z^i \qquad q_0 \neq 0 \\ P_p(z) &= \sum_{i=0}^{p} p_i z^i \qquad p_0 = 1 \; . \end{aligned} \tag{1.5}$$

then we say that the process $\{X_t\}$ satisfies a finite parameter autoregressive moving average model or simply ARMA(p,q). Usually we write

$$P_p(B)X_t = Q_q(B)e_t \tag{1.6}$$

where B is the backshift operator. There are two related problems to be considered here. The first is to determine the orders of the polynomials $P_p(z)$ and $Q_q(z)$. This is the model identification problem. The second is the problem of estimating the coefficients in $P_p(z)$ and $Q_q(z)$ after the model is identified. Given a model of the form (1.6), most of the literature assumes that $P_p(z)$ and $Q_q(z)$ have no root on the unit disk, $|z| \leq 1$. The condition $P_p(z) \neq 0$ for all $|z| \leq 1$ is called the realizability condition so that X_t has a one-sided infinite moving average representation

$$X_t = A(B)e_t = \sum_{j=0}^{\infty} a_j e_{t-j} \tag{1.7}$$

with

$$A(B) = \sum_{j=0}^{\infty} a_j B^j .$$

This is the same as saying $a_j = 0$ for all $j < 0$ in (1.1). The condition $Q_q(z) \neq 0$, $|z| \leq 1$, called the invertibility condition, is not needed for stationarity. If $\{X_t\}$ satisfies (1.6) and is a Gaussian process then it is well known that any real root $r_j \neq 0$ of $P_p(z)$ or $Q_q(z)$ can be replaced by its inverse r_j^{-1} and paired conjugate complex roots can be replaced by their conjugated inverses $\bar{r}_j^{-1}$ without changing the correlation structure of $\{X_t\}$. This means that if all the roots are real and distinct then there are 2^{p+q} different ways to specify the roots and they are indistinguishable by examining the autocorrelation function. Since different set of roots correspond to

different set of coefficients, it is customary to assume that all roots of $P_p(z)$ and $Q_q(z)$ are outside the unit circle and to estimate the coefficients of $P_p(z)$ and $Q_q(z)$ under this condition. We will present a method that can be applied without imposing the invertibility assumption.

There are various procedures in the literature concerning the identification of the orders p of $P_p(z)$ and q of $Q_q(z)$. Most of these procedures involve the examination of the residuals or estimates of e_t's. In doing so, invertibility is assumed. The distribution of e_t is assumed to be Gaussian or a known one so that maximum likelihood estimation of the coefficients can be carried out. Box and Jenkins [1976] considered an iterative procedure by examining the autocorrelation function and partial autocorrelation function. In a series of papers, Akaike [1969, 1971, 1978] proposed a final prediction error criteria (FPE), an information criteria (AIC) and a Bayesian version of it (BIC). These methods are based on multiple decision procedure and were studied by others. (See Priestley [1981]). Hypothesis testing methods were considered by Godfrey [1979] and Poskitt and Tremayne [1980]. Gray, Kelly and Woodward [1978] considered the S Array method using a pattern recognition technique. More recently Woodward and Gray [1981] proposed a generalized partial autocorrelation method. Tiao and Tsay [1981] proposed an iterative regression approach based on extended autocorrelation function. Hannan and Rissanen [1982] considered a recursive method to identify an ARMA model.

Parameter estimation methods have been developed by Hannan [1969], Box and Jenkins [1976], Anderson [1977] and Ansley [1979].

If the process $\{X_t\}$ is nonGaussian, Lii and Rosenblatt [1982] proved that, under broad conditions, (1.2) is identifiable up to a sign change and/or index shift of the a_j's requiring only that $P_p(z)$ and $Q_q(z)$ have no root of absolute value one.

It has been observed that in many geophysical and economic context that data is often nonGaussian. In this paper we propose a method to identify the orders p and q of the model (1.6) and estimate the corresponding coefficients without the usual invertibility assumption. In section 2 we adopt the higher order spectrum method proposed in Rosenblatt [1980] and Lii and Rosenblatt [1982] to estimate the a_j's in (1.7) and obtain their asymptotic distributions. In section 3 we introduce C-table and the Pade table and give a method to identify the model and to estimate the underlying parameters. Asymptotic results are given. Section 4 consists of a few examples and a discussion.

2. Asymptotics of the higher order spectral method

Let the frequency response function from (1.2) be

$$A(e^{-i\lambda}) = \left|2\pi f(\lambda)\right|^{\frac{1}{2}} \exp\{ih(\lambda)\} . \tag{2.1}$$

There are many references concerning the estimation of the spectral density function $f(\lambda)$. (See Anderson [1971] or Jenkins and Watts [1968]). Lii and Rosenblatt [1982] proposed a method to estimate the phase information $h(\lambda)$ when the process X_t (and hence the innovation process e_t) is nonGaussian. Some basic results from this paper are summarized in the following lemmas.

<u>Lemma 2.1.</u> Let $\{X_t\}$ be a nonGaussian linear process given in (1.1) with the independent random variable $\{e_t\}$ having all moments finite. Assuming that

$$\Sigma|j|\ |a_j| < \infty$$

and

$$A(e^{-i\lambda}) \neq 0 \quad \text{for all } \lambda$$

and

$$h(0) = 0 \tag{2.2}$$

Then the phase $h(\lambda)$ in (2.1) is given by

$$h(\lambda) = h_1(\lambda) - \lambda h_1(\pi)/\pi + a\lambda \tag{2.3}$$

where a is an integer and

$$h_1(\lambda) = \int_0^\lambda (h'(u) - h'(0))du \tag{2.4}$$

with

$$h'(0) - h'(\lambda) = \lim_{\Delta\to 0} \frac{1}{(m-2)\Delta} \{h(\lambda)+(m-2)h(\Delta)-h(\lambda+(m-2)\Delta)\} \tag{2.5}$$

where $m>2$ is an integer such that C_m, the m^{th} order cumulant of $\{X_t\}$, is nonzero and

$$h(\lambda_1) +\ldots+ h(\lambda_{m-1}) - h(\lambda_1 +\ldots+ \lambda_{m-1})$$

$$= \arg\{(\frac{A(1)}{|A(1)|})^m C_m^{-1} b(\lambda_1,\ldots,\lambda_{m-1})\} \tag{2.6}$$

where $b(\cdot)$ is the m^{th} order cummulant spectral density of the process $\{X_t\}$ discussed in Brillinger and Rosenblatt (1967).

Remark 1. From (1.2) and (2.1) we have

$$A(1) = \Sigma a_j = |2\pi f(0)|^{\frac{1}{2}} \exp\{ih(0)\} .$$

Since a_j's are real, we have either $\Sigma a_j > 0$ or $\Sigma a_j < 0$. The assumption $h(0) = 0$ in Lemma 2.1 represents an arbitrary choice of the signs

of a_j's. Observing X_t's only, the signs of the a_j's are intrinsically undecidable since we can multiply all a_j's and e_t's by minus one without changing (1.1).

The integer a in (2.3) is intrinsically undecidable also since it corresponds to reindexing the X_t's.

Remark 2. However, in the usual normalization of model (1.6) or (1.7) we assume $a_0 > 0$. Under this assumption we can use Theorem 2.1 to be proved later to ascertain the first nonzero a_j and shift the index and adjust the sign accordingly. Without loss of generality, in what follows, we will assume, that $m = 3$ in order to illustrate the techniques of the method.

<u>Lemma 2.2.</u> Under the assumptions of Lemma 2.1. An estimate of $h_1(\lambda)$ is, from (2.4 - 2.6),

$$H_n(\lambda) = -\sum_{j=1}^{k-1} \arg b_n(j\Delta,\Delta) \tag{2.7}$$

where $k\Delta = \lambda$ and it is understood that the bispectral estimates $b_n(\cdot)$ based on a sample of size n are weighted averages of third order periodogram values. If $b(\lambda,\mu) \in C^2$ and the weight function W is symmetric and band limited with band width Δ, then

$$H_n(\lambda) - h_1(\lambda) = R_n(\lambda) + o_p(H_n(\lambda) - h_1(\lambda))$$

with

$$ER_n(\lambda) = \Delta G(\lambda) + o(\Delta)$$

and

$$\begin{aligned} \mathrm{Cov}(R_n(\lambda), R\,(\mu)) &= \frac{\pi f(0)}{\Delta^3 n} \int_0^{\min(\lambda,\mu)} \frac{f^2(u)}{|b(u,0)|^2}\, du \cdot \\ &\qquad \int W^2(u,v)\,du\,dv \\ &= \frac{2\pi^2}{\Delta^3 n C_3^2} \min(\lambda,\mu) \int W^2(u,v)\,du\,dv \end{aligned} \tag{2.8}$$

for $\Delta(n) \to 0$, $\Delta^2 n \to \infty$ as $n \to \infty$

where $G(\lambda)$ is a function involving $b(\lambda,\mu)$. Further $EH_n(\lambda_j) \to h_1(\lambda_j)$ and $H_n(\lambda_j)$'s are asymptotically jointly normally distributed with covariances given by (2.8).

Since

$$a_j = \frac{1}{2\pi}\int_0^{2\pi} A(e^{-i\lambda})e^{ij\lambda}d\lambda$$

an estimate $\hat{a}_j$ of a_j is given by

$$\hat{a}_j = \frac{1}{2\pi}\int_0^{2\pi} A(e^{-i\lambda})e^{ij\lambda}\, d\lambda$$

$$\simeq \frac{1}{M}\sum_{k=1}^{M} (2\pi f_n(\lambda_k))^{\frac{1}{2}} \exp\left\{i(H_n(\lambda_k) - \frac{H_n(\pi)}{\pi}\lambda_k + j\lambda_k)\right\} \qquad (2.9)$$

which by symmetry can be written as

$$= \frac{2}{M}\sum_{k=1}^{L} (2\pi f_n(\lambda_k))^{\frac{1}{2}} \operatorname{Cos}\left(H_n(\lambda_k) - \frac{H_n(\pi)}{\pi}\lambda_k + j\lambda_k\right)$$

where $2L = M = 2\pi/\Delta$ and λ_k represent a discretization and f_n is an estimate of $f(\cdot)$ similar to that of $b(\cdot)$. For a given sample of size n, let the bandwidth of the weight function W_1 in $f_n(\lambda)$ be Δ_1 and the bandwidth of the weight function W_2 in $b_n(\lambda,\mu)$ be Δ_2, we will derive the asymptotic joint distribution of the $\hat{a}_j$'s given in (2.9) to the first order.

It is proved in Brillinger and Rosenblatt (1967) that if for $i=1,2$, $\Delta_i \to 0$ and $n\Delta_i^3 \to \infty$ as $n \to \infty$, then asymptotically as $n \to \infty$ $f_n(\lambda_k)$ and $b_n(\lambda_j,\mu_i)$ are independent normally distributed with

$$\operatorname{var}(f_n(\lambda_k)) = \frac{2\pi}{\Delta_1 n} f^2(\lambda_k) \int W_1^2(u)du \qquad \lambda_k \neq 0,\pi \;.$$

Since H_n is a function of b_n, hence H_n and f_n are asymptotically independent. Let

$$d_{\ell,k} = f_n^{\frac{1}{2}}(\lambda_\ell)\mathrm{Cos}(Z_n(\lambda_\ell)+k\lambda_\ell)$$

with

$$Z_n(\lambda_\ell) = H_n(\lambda_\ell) - \frac{H_n(\pi)}{\pi}\lambda_\ell$$

and observe that the asymptotic distribution of the vector $(f_n(\lambda_\ell),$ $H_n(\lambda_\ell),\ f_n(\lambda_j),\ H_n(\lambda_j),\ H_n(\pi))$ is normal with mean $(f(\lambda_\ell),\ h_1(\lambda_\ell),$ $f(\lambda_\ell),\ h_1(\lambda_j),\ h_1(\pi))$ and covariance matrix

$$\begin{bmatrix} s_{\ell,\ell} & 0 & s_{\ell,j} & 0 & 0 \\ 0 & t_{\ell,\ell} & 0 & t_{\ell,j} & t_{\ell,\pi} \\ s_{j,\ell} & 0 & s_{j,j} & 0 & 0 \\ 0 & t_{j,\ell} & 0 & t_{j,j} & t_{j,\pi} \\ 0 & t_{\pi,\ell} & 0 & t_{\pi,j} & t_{\pi,\pi} \end{bmatrix}$$

where

$$s_{\ell,j} = \mathrm{Cov}(f_n(\lambda_\ell),f_n(\lambda_j))$$

$$t_{\ell,j} = \mathrm{Cov}(H_n(\lambda_\ell),H_n(\lambda_j)) \text{ with } \lambda_\pi = \pi$$

we note that the magnitude of $s_{\ell,j}$ is smaller than that of $t_{\ell,j}$. An application of a multivariate δ-method (see Bishop, Fienberg and Holland [1975] p. 493) we can show that the asymptotic distribution of $(d_{\ell,k},d_{j,m})$ is bivariate normal with mean

$$(f^{\frac{1}{2}}(\lambda_\ell)\mathrm{Cos}(Z(\lambda_\ell)+k\lambda_\ell),f^{\frac{1}{2}}(\lambda_j)\mathrm{Cos}(Z(\lambda_j)+m\lambda_j))$$

where

$$Z(\lambda_\ell) = h_1(\lambda_\ell) - \lambda_\ell h_1(\pi)/\pi$$

and covariance matrix

with

$$K \begin{bmatrix} C(\ell,k;\ell,k) & C(\ell,k;j,m) \\ C(j,m;\ell,k) & C(j,m;j,m) \end{bmatrix}$$

$$C(\ell,k;j,m) = f^{\frac{1}{2}}(\lambda_\ell) f^{\frac{1}{2}}(\lambda_j) \operatorname{Sin}(Z(\lambda_\ell)+k\lambda_\ell) \cdot$$

$$\operatorname{Sin}(Z(\lambda_j)+m\lambda_j)\{\min(\lambda_\ell,\lambda_j) - \lambda_\ell\lambda_j/\pi\} \qquad (2.10)$$

and

$$K = \frac{2\pi^2}{\Delta_2^3 n C_3^2} \int W_2^2(u,v)dudv$$

Using this and a straightforward calculation, we have the following theorem.

<u>Theorem 2.1</u>. Under the assumption of Lemma 2.1 we have $(\hat{a}_k - a_k)$ for $k=1,\ldots,K$ are asymptotically jointly Gaussian with means zero and covariances given by

$$\operatorname{Cov}(\hat{a}_k,\hat{a}_m) = \frac{2(2\pi)^3}{M^2\Delta_2^3 n C_3^2} \int W_2^2(u,v)dudv \sum_{\ell=1}^{L} \sum_{j=1}^{L} C(\ell,k;j,m) \qquad (2.11)$$

where $C(\ell,k;j,m)$ is given in (2.10)

We will now assume that the stationary process $\{X_t\}$ satisfies (1.6) with a representation given in (1.7) such that $a_0 > 0$. As usual, we assume $P_p(z)$ and $Q_q(z)$ given in (1.5) have no common factor with $p_0 = 1$ and $q_0 > 0$. Under these assumptions, equation (2.11) can be used to estimate the variance of $\hat{a}_k$ with $f(\lambda)$ and $h_1(\lambda)$ estimated by $f_n(\lambda)$ and $H_n(\lambda_\ell)$ respectively. These results can be used to determine the smallest integer k such that $E\hat{a}_k \neq 0$. We then reindex the a_j's and change their signs if necessary. This gives a complete procedure to estimate a_j's in equation (1.7) consistently. We use these estimated $\hat{a}_j$'s to identify and estimate the polynomials $P_p(z)$ and $Q_q(z)$ by the C-table and the Pade table as discussed in Lii [1982]

dealing with a distributed lag model.

3. Asymptotics of the C-table and the Pade approximant

Given a pair of nonnegative integers q and p we denote Pade rational approximants to a formal power series $A(z) = \sum_{j=0}^{\infty} a_j z^j$ by $[q/p] = Q_q(z)/P_p(z)$ where $Q_q(z)$ and $P_p(z)$ are polynomials of degrees at most q and p respectively. We assume $P_p(0) = 1$ and $Q_q(z)$ and $P_p(z)$ have no common factors. The coefficients of $Q_q(z)$ and $P_p(z)$ are determined by $A(z) - (Q_q(z)/P_p(z)) = 0(z^{p+q+1})$. The following three lemmas can be found in Baker [1975].

Lemma 3.1. When it exists, the $[q/p]$ Pade approximant for A(z) is uniquely determined. Further

$$Q_q(z) = \det \begin{vmatrix} a_{q-p+1} & a_{q-p+2} & \cdots & a_{q+1} \\ a_{q-p+2} & a_{q-p+3} & \cdots & a_{q+2} \\ \vdots & \vdots & & \vdots \\ a_q & a_{q+1} & \cdots & a_{q+p} \\ \sum_{j=p}^{q} a_{j-p} z^j & \sum_{j=p-1}^{q} a_{j-p+1} z^j & \cdots & \sum_{j=0}^{q} a_j z^j \end{vmatrix} \tag{3.1}$$

and

$$P_p(z) = \det \begin{vmatrix} a_{q-p+1} & a_{q-p+2} & \cdots & a_{q+1} \\ a_{q-p+2} & a_{q-p+3} & \cdots & a_{q+2} \\ \vdots & \vdots & & \vdots \\ a_q & a_{q+1} & & a_{q+p} \\ z^p & z^{p-1} & \cdots & 1 \end{vmatrix} \tag{3.2}$$

where $a_j = 0$ if $j < 0$ and the summation is set to zero if the lower index on a sum exceeds the upper index.

Given nonnegative integers r and s, we define

$$C_{r,s} = \det \begin{vmatrix} a_{r-s+1} & a_{r-s+2} & \cdots & a_r \\ a_{r-s+2} & a_{r-s+3} & \cdots & a_{r+1} \\ \vdots & \vdots & & \vdots \\ a_r & a_{r+1} & & a_{r+s-1} \end{vmatrix} \tag{3.3}$$

$$= (-1)^{\frac{s(s-1)}{2}} \det\left[(a_{r+i-j})_{i,j=1}^{s}\right] .$$

The C-table,which is a doubly infinite array, is defined by $C = (C_{r,s})_{r,s=0}^{\infty}$. We further define $C_{r,0} = 1$.

<u>Lemma 3.2</u>. (i) $C_{q,p} \neq 0$ implies that $[q/p]$ exists. (ii) Every zero entry in the C-table for a formal power series $A(z) = 1 + \sum_{j=1}^{\infty} a_j z^j$ occurs in a square block of zero entries and is completely boarded by nonzero entries.

<u>Lemma 3.3</u>. Given a formal power series A(z) the following three conditions are equivalent

(1) $A(z) = \sum_{j=0}^{\ell} c_j z^j / (1 + \sum_{j=1}^{m} d_j z^j)$

(2) $[q/p] = A(z)$ for all $q \geq \ell$ and $p \geq m$

(3) $C_{q,p} \neq 0$ and $C_{r,s} = 0$ for all $r > \ell$ and $s > m$.

If condition (3) in Lemma 3.3 is satisfied, we call the entry $(\ell+1, m+1)$ in the C-table the "breaking point".

Lemmas 3.1, 3.2 and 3.3 lead to the following.

<u>Theorem 3.1</u>. The process $\{X_t\}$ given in (1.7) is ARMA (p,q) given in (1.6) if and only if the C-table associated with A(z) has the breaking point (p+1,q+1). Further $C_{q,p} \neq 0$ and the coefficients of $P_p(z)$ and $Q_q(z)$ in (1.6) are obtained from (3.1) and (3.2). To normalize these

coefficients we divided both $P_p(z)$ and $Q_q(z)$ by $C_{q,p}$ so that $p_0 = 1$. Whether the roots of $Q_q(z)$ are inside or outside of the unit circle is immaterial here.

This theorem provides a consistent procedure to identify the model by determining the orders p and q and to estimate the parameters of the identified model. We use the $\hat{a}_j$'s obtained from (2.9) to construct estimates $\hat{C}_{r,s}$ of $C_{r,s}$, ascertain the breaking point in $\hat{C}$-table to identify the model and finally by substituting the $\hat{a}_j$'s for a_j's in (3.1) and (3.2) to obtain estimates of the coefficients of $P_p(z)$ and $Q_q(z)$. The following lemmas can now be proved with simple modifications of the proofs given in section 4 of Lii [1982].

<u>Lemma 3.4</u>. If $\{\hat{a}_j\}_{j=1}^{K}$ are asymptotically jointly Gaussian for a fixed integer K with mean $\{a_j\}_{j=1}^{K}$ and covariance matrix $\Sigma(C_n)$ where $C_n = n^{-\delta} \to 0$ $\delta > 0$ as the sample size $n \to \infty$ then the asymptotic distribution of $\hat{R}$, the determinant of an M×M matrix,

$$\hat{R} = \det\left[(\hat{a}_{i,j})_{i,j=1}^{M}\right] \text{ with } \hat{a}_{i,j} \in \{\hat{a}_\ell\}_{\ell=1}^{K}, \text{ is Gaussian with mean}$$

$$R = \det\left[(a_{i,j})_{i,j=1}^{M}\right]$$

and variance

$$\sigma_R^2 = G\Sigma(C_n)G^t \text{ where } G^t \text{ is the transpose of } G = (g_1,\ldots,g_K) \text{ with}$$

$$g_j = \frac{\partial}{\partial a_j} R.$$

We note that computationally g_j is the sum of the cofactors of a_j in the matrix $[(a_{i,j})_{i,j=1}^{M}]$.

<u>Theorem 3.2</u>. If the estimates of $\{a_j\}$ in (1.7) are given by $\{a_j\}$ obtained from (2.9), then for fixed r and s,

$$\hat{C}_{r,s} = (-1)^{s(s-1)/2} \det \left[(\hat{a}_{r+i-j})_{i,j=1}^{s}\right]$$

is asymptotically normally distributed with mean $C_{r,s}$ given in (3.3) and variance $G\Sigma G'$ where $G = (g_L, g_{L+1}, \ldots, g_U)$ with $L = r-s+1$, $U = r+s-1$, $g_j = \frac{\partial}{\partial a_j} C_{r,s}$ and Σ is the covariance matrix of $(\hat{a}_L, \ldots, \hat{a}_U)$ from (2.11).

This theorem gives a method to construct the $\hat{C}$-table and to find the breaking point $(q+1, p+1)$. If the breaking point can not be uniquely determined, the $\hat{C}$-table will reduce the number of possible competing models to only a few for further testing. If the process $\{X_t\}$ does not have a rational frequency response function $[q/p]$, the $\hat{C}$-table will still suggest a possible ARMA (p,q) approximation to A(z) using the principle of parsimony. Once we have identified the model to be ARMA (p,q), replacing the a_j's by their estimates $\hat{a}_j$'s in equations (3.2), we obtain estimates $\hat{p}_i$ and $\hat{q}_i$ of p_i and q_i respectively by

$$\frac{\hat{Q}_q(z)}{\hat{P}_p(z)} = \frac{\hat{q}_0'+\hat{q}_1 z + \ldots + \hat{q}_q' z^q}{\hat{p}_0'+\hat{p}_1' z + \ldots + \hat{p}_p' z^p}$$

$$= \frac{\hat{q}_0+\hat{q}_1 z + \ldots + \hat{q}_q z^q}{1 + \hat{p}_1 z + \ldots + \hat{p}_p z^p} \tag{3.4}$$

with

$$\hat{p}_0' = \hat{C}_{q,p} \neq 0$$

$$\hat{p}_i = \hat{p}_i'/\hat{p}_0' \qquad i = 0,1,\ldots,p$$

$$\hat{q}_i = \hat{q}_i'/\hat{p}_0' \qquad i = 0,1,\ldots,q\ . \tag{3.5}$$

To obtain the asymptotic distributions of $\hat{p}_i$'s and $\hat{q}_i$'s we evaluate

the determinant in (3.2) by cofactor expansion of the last row $(z^p, z^{p-1}, \ldots, 1)$ and we obtain

$$\hat{P}_p(z) = \hat{A}_0 - \hat{A}_1 z + \hat{A}_2 z^2 + \ldots + (-1)^p \hat{A}_p z^p \tag{3.6}$$

where $\hat{A}_i$ is the cofactor of Z^i in (3.2).

Similarly, we have from (3.1)

$$\begin{aligned}\hat{Q}_q(z) &= \hat{A}_0\Big(\sum_{j=0}^{q} \hat{a}_j z^j\Big) - \hat{A}_1\Big(\sum_{j=2}^{q} \hat{a}_{j-1} z^j\Big) + \ldots + (-1)^p \hat{A}_p\Big(\sum_{j=p}^{q} \hat{a}_{j-p} z^j\Big) \\ &= \hat{a}_0\hat{A}_0 + (\hat{a}_1\hat{A}_0 - \hat{a}_0\hat{A}_1)z + \ldots + (\hat{a}_q\hat{A}_0 + \ldots + (-1)^p \hat{a}_{q-p}\hat{A}_p)z^q \\ &= \hat{B}_0 + \hat{B}_1 z + \ldots + \hat{B}_q z^q \end{aligned} \tag{3.7}$$

Using Lemma 3.4 and equations (3.4) - (3.7) we can prove

<u>Theorem 3.3</u>. For fixed p and q, let $L = \max\{0, q-p+1\}$ and $U = q+p-1$. Then the asymptotic distributions of $(\hat{p}_i' - p_i', \hat{p}_j' - p_j')$, $(\hat{p}_i' - p_i', \hat{q}_j' - q_j')$ and $(\hat{q}_i' - q_i', \hat{q}_j' - q_j')$ are each bivariate normal with mean (0,0) and covariance matrices

$$\Sigma_{i,j}^{PP} = \begin{pmatrix} P_{i,L,U} \\ P_{j,L,U} \end{pmatrix} \Sigma_{L,U} (P_{i,L,U}^t, P_{j,L,U}^t)$$

$$\Sigma_{i,j}^{PQ} = \begin{pmatrix} P_{i,L,U} \\ Q_{j,L,U} \end{pmatrix} \Sigma_{L,U} (P_{i,L,U}^t, Q_{j,L,U}^t)$$

and

$$\Sigma_{i,j}^{QQ} = \begin{pmatrix} Q_{i,L,U} \\ Q_{j,L,U} \end{pmatrix} \Sigma_{L,U} (Q_{i,L,U}^t, Q_{j,L,U}^t)$$

where

$$P_{i,L,U} = (g_L, g_{L+1}, \ldots, g_U) \quad i = 0, \ldots, p$$

$$Q_{j,L,U} = (h_L, h_{L+1}, \ldots, h_U) \quad j = 0, \ldots, q$$

with

$$g_\ell = \frac{\partial}{\partial a_\ell} A_\ell \quad \ell=L,\ldots,U$$

$$h_\ell = \frac{\partial}{\partial a_\ell} B_\ell \quad \ell=L,\ldots,U$$

and

$$\Sigma_{L,U} = \mathrm{Cov}(\hat{a}_L,\ldots,\hat{a}_U) \quad \text{from (2.11).}$$

A_ℓ and B_ℓ are the theoretical values of $\hat{A}_\ell$ and $\hat{B}_\ell$ respectively in (3.6) and (3.7). Furthermore, the asymptotic distribution of $\hat{p}_i - p_i$ and $\hat{q}_j - q_j$ are normal with mean zero and variances

$$\sigma^2_{p_i} = \left(\frac{1}{p_0'}, -\frac{p_i'}{p_0'^2}\right) \Sigma^{PP}_{0i} \begin{pmatrix} \frac{1}{p_0'} \\ -\frac{p_i'}{p_0'^2} \end{pmatrix}$$

$$\sigma^2_{q_j} = \left(\frac{1}{p_0'}, -\frac{q_j'}{p_0'^2}\right) \Sigma^{PQ}_{0j} \begin{pmatrix} \frac{1}{p_0'} \\ -\frac{q_j'}{p_0'^2} \end{pmatrix} .$$

4. Examples and Discussion

Examples in this section are simulated according to the model of the form $P_p(B)X_t = Q_q(B)e_t$ with

$$P_p(B) = 1 + p_1 B + \ldots + p_p B^p$$

$$Q_q(B) = q_0 + q_1 B + \ldots + q_q B^q \quad q_0 > 0$$

and

$$P_p(z) \neq 0 \quad \text{when } |z| \leq 1 .$$

The innovation process are obtained from

$$e_t = \frac{1}{\sigma}(e_t' - \mu)$$

where e_t' are independent, identical, exponentially distributed with $\mu = Ee_t' = 1$ and $\sigma^2 = Var(e_t') = 1$. Hence $Ee_t = 0$, $Var(e_t) = 1$. The sample size for X_t is 640. Some computational details are discussed in Lii and Helland [1981] and Lii and Rosenblatt [1982].

Example 1.

$$Q_2(B) = 1 - 0.6B + 0.8B^2$$

$$P_1(B) = 1 + 0.6B$$

All the roots are outside of unit circle. Table 1 gives the $\hat{C}$-table associated with this model. Each entry has two numbers, the upper one is $\hat{C}_{r,s}$ and the lower one is the estimated standard deviation of $\hat{C}_{r,s}$ computed from Theorem 3.2. We also exhibit table 2 which gives the ratio of the $\hat{C}_{r,s}$ and its estimated standard deviation of each entry in table 1. We call table 2 the "resolution table" of table 1. It is much easier to recognize the pattern in a resolution table when there is a sudden drop of resolution at entry (ℓ,m) and thereafter (ℓ,m) is likely to be the breaking point. From table 2, it is clear that (3,2) is the breaking point and the model is correctly identified as ARMA (2,1). The Pade approximant [2/1] gives, from (3.4) and (3.5),

$$Q_2(B) = \underset{(0.446)}{1.068} - \underset{(0.212)}{0.585}\ B^2 + \underset{(0.097)}{0.763}\ B^2$$

and

$$P_1(B) = 1 + \underset{(0.092)}{0.594}\ B$$

where the numbers in the parentheses are estimated standard deviations from theorem 3.3.

Example 2. In this example both roots, -0.5 and -0.75, of

$$Q_2(B) = 1 + 3.5B + 3B^2$$

are inside of the unit circle while the roots of $P_2(B) = 1 + 0.35B + 0.5B$ are outside of unit circle. The associated $\hat{C}$-table is table 3 and its resolution table is table 4. It seems reasonable to identify the model to be ARMA(2,2) with breaking point at (3,3). The Pade approximant [2/2] gives

$$\begin{aligned} Q_2(B) &= \underset{(0.94)}{0.907} + \underset{(13.7)}{3.64}\, B + \underset{(10.5)}{2.86}\, B^2 \\ \text{and} \qquad & \qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.1) \\ P_2(B) &= 1 + \underset{(0.60)}{0.55}\, B + \underset{(0.62)}{0.53}\, B^2 \end{aligned}$$

The large estimated standard deviations in Example 2 may be due to the complicated formula in Theorem 3.3 and the number of parameters are large relative to the sample size. Nevertheless, the estimates of the parameters provide good starting values for possible more efficient iterative methods. We note that the usual iterative type of fitting procedure can be used here. We can deconvolve the process X_t and estimate the innovation process e_t by $\hat{e}_t$. Diagnostic checking can be performed on $\hat{e}_t$ to discriminate among possible competing models. The probability distribution or density function of e_t can be estimated to facilitate a non-Gaussian maximum likelihood estimation. It seems that in building a finite parameter ARMA model of a stationary time series $\{X_t\}$, one should use the procedure suggested in Lii and Rosenblatt [1982] to deconvolve X_t and see if $\{e_t\}$ is near Gaussian or not. If not, one should use the procedure suggested in this paper to build the ARMA model without imposing the invertibility condition. Alternatively, one may want to use any one of those methods mentioned in the introduction section, using mainly the second order structure, to identify the orders of the model; however one should still use the Pade approximant to estimate necessary coefficients and to identify

whether the roots lie inside or outside the unit circle. Even in the Gaussian case, one may want to first fit an MA(K) for a moderate integer K (say 15). Then following the procedure in section 3, one can identify the equivalent parsimonious ARMA(p,q) model and obtain estimates of parameters. As a comparison, we employed the usual Box-Jenkins type estimation procedure as it is implemented in the subroutine FTML of the International Mathematical and Statistical Library (ISML). Given the right orders in the model, we obtain estimates

$$Q_2(B) = 1.0 + 1.161B + 0.3027B^2$$

and (4.2)

$$P_2(B) = 1.0 + 0.3307B + 0.4789B^2$$

with estimated white noise variance $\sigma_e^2 = 8.765$.

Using $Q_2(B)$ in (4.2) to interprete the model may be quite different from that of using $Q_2(B)$ in (4.1).

Example 2 shows that we can discriminate models which are indistinguishable using only second order properties. The method proposal in this paper produce estimates that are consistent. For moderate sample size this method can be a valuable tool for ARMA model identification and estimation.

Table 1

	1	2	3	4	5	6	7	8
0	0.107E+01 0.195E+00	-0.114E+01 0.417E+00	-0.122E+01 0.668E+00	0.130E+01 0.952E+00	0.139E+01 0.127E+01	-0.149E+01 0.163E+01	-0.159E+01 0.203E+01	0.170E+01 0.249E+01
1	-0.122E+01 0.107E+00	0.101E+00 0.750E-01	-0.105E+01 0.292E+00	-0.493E+00 0.269E+00	-0.511E+00 0.188E+00	-0.460E+00 0.266E+00	-0.256E+00 0.160E+00	-0.471E+00 0.346E+00
2	0.149E+01 0.470E-01	-0.114E+01 0.993E-01	-0.870E+00 0.134E+00	0.599E+00 0.114E+00	0.350E+00 0.991E-01	-0.230E+00 0.797E-01	-0.177E+00 0.630E-01	0.136E+00 0.502E-01
3	-0.885E+00 0.155E+00	-0.507E-01 0.108E+00	-0.731E-01 0.627E-01	-0.110E+00 0.401E-01	0.304E-01 0.199E-01	0.198E-01 0.136E-01	-0.952E-03 0.115E-01	-0.155E+00 0.583E-01
4	0.492E+00 0.142E+00	-0.592E-01 0.511E-01	-0.126E-01 0.170E-01	0.166E-01 0.107E-01	-0.362E-02 0.294E-02	-0.158E-02 0.214E-02	0.173E-01 0.124E-01	0.174E+00 0.593E-01
5	-0.207E+00 0.102E+00	0.528E-01 0.206E-01	0.113E-01 0.733E-02	-0.292E-02 0.235E-02	-0.435E-03 0.787E-03	-0.304E-02 0.254E-02	-0.250E-01 0.148E-01	-0.265E+00 0.765E-01
6	0.194E+00 0.753E-01	-0.766E-02 0.900E-02	0.209E-02 0.168E-02	0.216E-03 0.582E-03	-0.250E-02 0.203E-02	-0.127E-01 0.825E-02	0.826E-01 0.272E-01	0.339E+00 0.151E+00
7	-0.145E+00 0.574E-01	0.881E-02 0.583E-02	0.241E-03 0.106E-02	0.177E-02 0.143E-02	-0.803E-02 0.526E-02	0.150E-01 0.149E-01	-0.101E+00 0.629E-01	-0.864E+00 0.240E+00

Table 2

	1	2	3	4	5	6	7	8
0	5.479	- 2.739	-1.826	1.369	1.095	0.000	0.000	0.000
1	-11.459	1.353	-3.609	-1.835	-2.715	-1.732	-1.597	-1.361
2	31.703	-11.442	-6.477	5.270	3.526	-2.884	-2.815	2.704
3	- 5.725	0.000	-1.165	-2.753	1.523	1.458	0.000	-2.654
4	3.476	- 1.158	0.000	1.549	-1.234	0.000	1.397	2.932
5	- 2.018	2.569	1.543	-1.243	0.000	-1.193	-1.686	-3.471
6	2.580	0.000	1.244	0.000	-1.228	-1.536	3.040	2.244
7	- 2.532	1.510	0.000	1.240	-1.525	1.005	-1.608	-3.590

Table 4

	1	2	3	4	5	6	7	8
0	2.204	-1.102	0.000	0.000	0.000	0.000	0.000	0.000
1	12.805	-7.448	-4.038	2.548	1.800	-1.372	-1.103	0.000
2	0.000	-4.057	-2.839	2.512	2.359	-2.167	-2.024	1.956
3	- 8.771	-4.102	0.000	0.000	0.000	-1.341	3.777	-1.841
4	2.436	-2.972	0.000	0.000	0.000	-1.143	-1.349	2.075
5	1.768	-2.731	0.000	0.000	0.000	0.000	0.000	-2.192
6	- 4.566	-1.857	1.344	0.000	0.000	0.000	0.000	2.526
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-2.082

References

1. Akaike, H. (1969). "Fitting autoregressive model for prediction." Annals of the Institute of Statistical Mathematics 21, 243-247.

2. Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle". In "2nd International Symposium on Information Theory", Ed. by B. N. Petrov and F. Caski, 267-281, Akademia Kiado, Budapest.

3. Akaike, H. (1978). "A Bayesian analysis of the minimum AIC procedure." Annals of the Institute of Statistical Mathematics 30, 9-14.

4. Anderson, T. W. (1971). "The spectral analysis of time series." John Wiley, New York.

5. Anderson, T. W. (1977). "Estimation for autoregressive moving average models in the time and frequency domains." The Annals of Statistics 6, 842-865.

6. Ansley, G. F. (1979). "An algorithm for the exact likelihood of a mixed autoregressive moving average process." Biometrica 66, 59-65.

7. Baker, G. A., Jr. (1975). "Essentials of Pade approximations." Academic Press, New York.

8. Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). "Discrete multivariate analysis - theory and practice." The MIT Press, Cambridge, Massachusetts.

9. Box, G. E. P., and Jenkins, G. M. (1976). "Time series analysis: forecasting and control." Holden-Day, San Francisco.

10. Brillinger, D. R. and Rosenblatt, M. (1967). "Asymptotic theory of estimates of kth order spectra." In "Spectral Analysis of Time Series", Ed. by B. Harris, 153-188. John Wiley, New York.

11. Chatfield, C. (1979). "Inverse autocorrelation." Journal of the Royal Statistical Society A 142, 363-377.

12. Godfrey, L. G. (1979). "Testing the adequacy of a time series model." Biometrica 66, 67-72.

13. Gray, H. L., Kelley, G. D., and McIntire, D. D. (1978). "A new approach to ARMA modeling." Communications in Statistics B7, 1-77.

14. Hannan, E. J. (1969). "The estimation of mixed moving average autoregressive system." Biometrica 56, 579-594.

15. Hannan, E. J. and Rissanen, J. (1982). "Recursive estimation of mixed autoregressive-moving average order." Biometrica 1, 81-94.

16. Jenkins, G. M. and Watts, D. G. (1968). "Spectral analysis and its application." Holden-Day, San Francisco.

17. Lii, K. S. and Helland, H. N. (1981). "Cross-bispectrum computation and variance estimation." ACM Transaction of Mathematical Software 7, 284-294.

18. Lii, K. S. and Rosenblatt, M. (1982). "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes." To appear in The Annals of Statistics, December.

19. Lii, K. S. (1982). "Model identification in a transfer function model." Technical report No. 95, Statistics Department, University of California, Riverside.

20. Poskitt, D. S. and Tremayne, A. R. (1980). "Testing the specification of a fitted autoregressive-moving average model." Biometrica 67, 359-363.

21. Priestley, M. B. (1981). "Spectral analysis and time series, Vol. 1." Academic Press, New York.

22. Rosenblatt, M. (1980). "Linear processes and Bispectra." Journal of Applied Probability 17, 265-270.

23. Tiao, G. C., and Tsay, R. S. (1981). "Identification of non-stationary and stationary ARMA models." In Proceedings of the Business and Economic Statistics Section. American Statistical Association, Washington, D.C.

24. Woodward, W. A., and Gray, H. L. (1981). "On the relationship between the S array and the Box-Jenkins method of ARMA model identification." Journal of the American Statistical Association 76, 579-587.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER --	2. GOVT ACCESSION NO. AD A -120662	3. RECIPIENT'S CATALOG NUMBER --
4. TITLE (and Subtitle) MODEL IDENTIFICATION AND ESTIMATION OF NONGAUSSIAN ARMA PROCESSES		5. TYPE OF REPORT & PERIOD COVERED Research 6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Keh-Shin Lii		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-81-K0003
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics University of California, San Diego La Jolla, California 92093		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS --
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE 1982 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) --		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unclassified.

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARMA model, identification, estimation, C-table, Pade table, Asymptotics, bispectrum

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Finite parameter models of ARMA type have been used extensively in many applications. Under the usual Gaussian assumption, the second order analysis will not be able to discriminate among competing models which give the same correlation structure. In many applications the innovation process is nonGaussian. In this case, analysis using higher order moments will identify the models uniquely without the usual invertibility assumption. This in turn will affect the forecasting based on the nonGaussian model. We present a method which uses bispectral analysis and the Pade approximation. We show that the (continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

method will consistently identify the order of the ARMA model and estimate the paramet ers of the model. One could also deconvolve the process to estimate the innovation process which will provide information for possible more efficient maximum likelihood estimation of the parameters. Asymptotic distributions are given, and a few examples are presented to illustrate the effectiveness of the method.